A Self-Organizing Neural Network For Supervised Learning, Recognition, and Prediction

Can neural networks learn to recognize new objects without forgetting familiar ones?

Gail A. Carpenter and Stephen Grossberg

As we humans move through our world, we can attend to both familiar and novel objects. Part of what makes us human is our ability to rapidly recognize, test hypotheses about, and name novel objects without disrupting our memories of familiar objects. This article describes a way of achieving these human characteristics in a self-organizing neural network called fuzzy ARTMAP. This architecture is capable of fast but stable on-line recognition learning, hypothesis testing, and adaptive naming in response to an arbitrary stream of analog or binary input patterns.

The fuzzy ARTMAP neural network combines a unique set of computational abilities that are needed to function autonomously in a changing world (see Table 1) and that alternative models have not yet achieved. In particular, fuzzy ARTMAP can autonomously learn, recognize, and make predictions about rare events, large nonstationary databases, morphologically variable types of events, and many-to-one and one-to-many relationships.

Fast Learning of Rare Events

An autonomous agent must be able to learn about rare events with important consequences, even if such events are similar to many other events that have different consequences (Fig. 1). For example, a rare medical case may be the harbinger of a new epidemic. A faint astronomical signal may signify important consequences for theories of the universe. A slightly different chemical assay may predict the biological effects of a new drug. Many traditional learning schemes use a form of slow learning that tends to average similar event occurrences. In contrast, fuzzy ARTMAP systems can rapidly learn rare events whose predictions differ from those of similar events.

Stable memory of Nonstationary Data

Rare events typically occur in a nonstationary environment, such as a large database, in which event statistics may change rapidly and unexpectedly. Individual events may also occur with variable frequencies and durations, and arbitrarily large numbers of events may need to be processed. Each of these factors tends to destabilize the learning process within traditional algorithms. New learning in such algorithms tends to unselectively wash away the memory traces of old, but still useful, knowledge. Learning a new face, for instance, could erase the memory of a parent's face, or learning a new type of expertise could erase the memory of previous expert knowledge.

A fuzzy ARTMAP system can reconcile conflicting properties and autonomously learn about:

Rare events
- requires fast learning

Large nonstationary databases
- requires stable learning

Morphologically variable events
- requires multiple scales of generalization (fine/coarse)

One-to-many and many-to-one relationships
- requires categorization, naming, and expert knowledge

To realize these properties, ARTMAP systems:

Pay attention
- ignore masses of irrelevant data

Test hypotheses
- discover predictive constraints hidden in data streams

Choose best answers
- quickly select globally optimal solution at any stage of learning

Calibrate confidence
- measure on-line how well a hypothesis matches the data

Discover rules
- identify transparent if-then relations at each learning stage

Scale
- preserve all desirable properties in arbitrarily large problems

■ **Table 1.** *Autonomous learning and control in a nonstationary world.*

GAIL CARPENTER is professor of Cognitive and Neural Systems (CNS) and Mathematics at Boston University.

STEPHEN GROSSBERG is Wang Professor of Cognitive and Neural Systems and Professor of Mathematics, Psychology, and Biomedical Engineering at Boston University.

0163-6804/92/$03.00 1992© IEEE

Adaptive Fitting of Morphological Variability

Many environments contain information that may be either coarsely or precisely defined. In other words, the morphological variability of the data may change through time. For example, it may be necessary merely to recognize that an object is an airplane, or that it is a particular type of airplane that is flown for a particular purpose by a particular country. Under autonomous learning conditions, no teacher is typically available to instruct a system about how coarse the definition of particular types of data should be. Multiple scales of generalization, from fine to coarse, need to be available on an as-needed basis. Fuzzy ARTMAP is able to automatically adjust its scale of generalization to match the morphological variability of the data. It embodies a Minimax Learning Rule that conjointly minimizes predictive error and maximizes generalization using only information that is locally available under incremental learning conditions in a nonstationary environment.

Learning Many-to-One and One-to-Many Maps

Autonomous agents must also be able to learn many-to-one and one-to-many relationships. Many-to-one learning takes two forms: categorization and naming (Fig. 2). For instance, during categorization of printed letter fonts, many similar samples of the same printed letter may establish a single recognition category, or compressed representation. Different printed letter fonts or written samples of the letter may establish additional categories. Each of these categories carries out a many-to-one map of its exemplars. During naming, all of the categories that represent the same letter may be associatively mapped into the letter name or prediction. There need be no relationship whatsoever between the visual features that define a printed letter A and a written letter A, yet both categories may need to be assigned the same name for cultural, not visual, reasons.

One-to-many learning is used to build up expert knowledge about an object or event (Fig. 3). A single visual image of a particular animal, for example, may lead to learning that predicts: animal, dog, beagle, and my dog Rover. Likewise, a computerized record of a patient's medical check-up may lead to a series of predictions about the patient's health; or a chemical assay of a sample of coal or petroleum may lead to many predictions about its uses as an energy source or material.

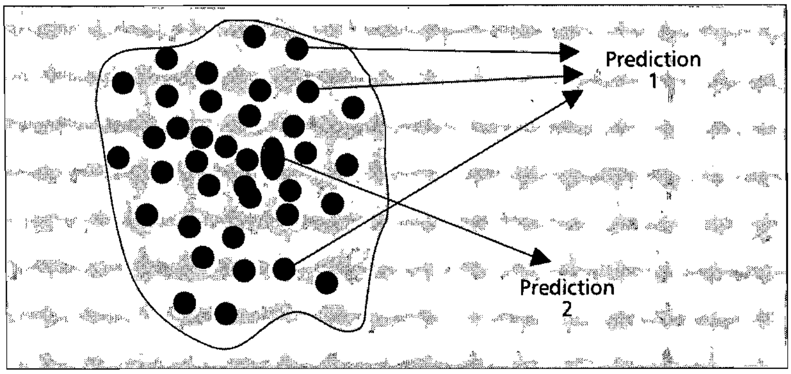
In many learning algorithms, the attempt to learn more than one prediction about an event leads to unselective forgetting of previously learned predictions, for the same reason that these algorithms become unstable in response to nonstationary data.

Error-Based Learning and Alternatives

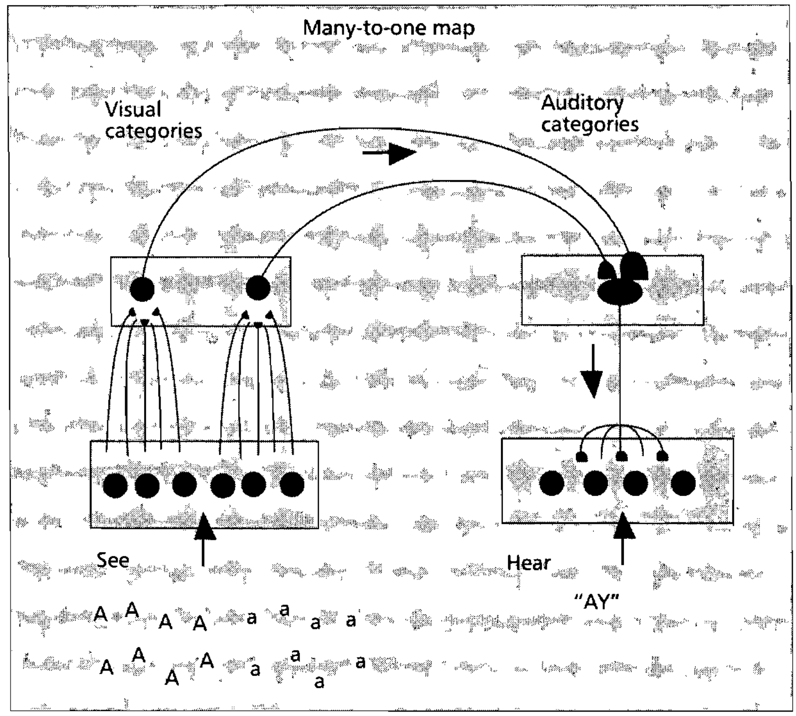
Error-based learning systems, including the back propagation algorithm, find it difficult, if not impossible, to achieve any of these computational goals [1-3]. Back propagation compares its actual prediction with a correct prediction and uses the error to change adaptive weights in a direction that is error-reducing. Fast learning would zero the error on each learning trial, and therefore cause massive forgetting. Statistical changes in the environment drag the adaptive weights away from their estimates of the previous environment. Longer event durations zero the error further, also destabilizing previous memories for the same reason that fast learning does. The selection of a fixed number of hidden units tends to fix a uniform level of generalization. Error-based learning also tends to force forgetting of previous predictions under one-to-many learning conditions, because the present correct prediction treats all previously learned predictions as errors. Ratcliff has noted, moreover, that back propagation fails to simulate human cognitive data about learning and forgetting [4].

Fuzzy ARTMAP exhibits the properties outlined so far in this article because it implements a qualitatively different set of heuristics than error-based learning systems. These heuristics are embodied in the following types of processes:

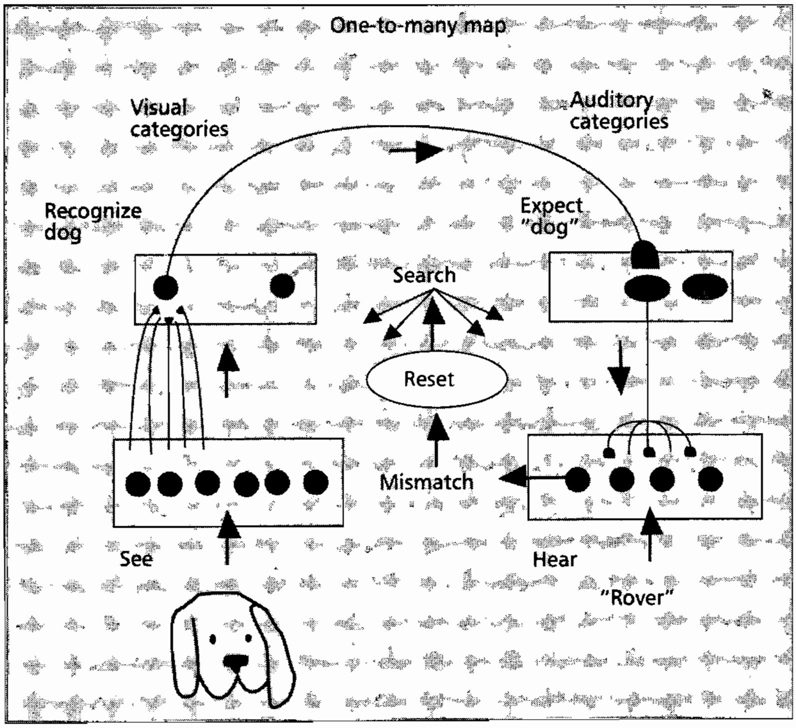
Pay attention—A fuzzy ARTMAP system can learn top-down expectations (also called primes, or queries) that enable the system to ignore masses of irrelevant



■ **Figure 1.** *Fuzzy ARTMAP can make a different prediction for a rare event than for all the similar events that surround it.*



■ **Figure 2.** *Many-to-one learning combines categorization of many exemplars into one category, and labeling of many categories with the same name.*



■ **Figure 3.** *One-to-many learning enables one input vector to be associated with many output vectors. If the system predicts an output that is disconfirmed at a given stage of learning, the predictive error drives a memory search for a new category to associate with the new prediction, without degrading its previous knowledge about the input vector.*

data. A large mismatch between a bottom-up input vector and a top-down expectation can drive an adaptive memory search that carries out hypothesis testing and match-based learning.

Carry out hypothesis testing and match-based learning—A fuzzy ARTMAP system actively searches for recognition categories, or hypotheses, whose top-down expectations provide an acceptable match to bottom-up data. The top-down expectation focuses attention upon and binds that cluster of input features that it deems to be relevant. If no available category or hypothesis provides a good enough match, then selection and learning of a new category and top-down expectation is automatically initiated. When the search discovers a category that provides an acceptable match, the system locks into an attentive resonance in which the input pattern refines the adaptive weights of the category based on any new information that it contains.

Thus, the fuzzy ARTMAP system carries out match-based learning, rather than error-based learning. A category modifies its previous learning only if its top-down expectation matches the input vector well enough to risk changing its defining characteristics. Otherwise, hypothesis testing selects a new category on which to base learning of a novel event.

Medical database - mortality following coronary bypass grafting (CABG) surgery
Fuzzy ARTMAP significantly outperforms:
- Logistic regression
- Additive model
- Bayesian assignment
- Cluster analysis
- Classification and regression trees
- Expert panel-derived sickness scores
- Principal component analysis

Mushroom database
- Decision trees (90-95% correct)
- ARTMAP (100% correct)
  Training set an order of magnitude smaller

Letter recognition database
- Genetic algorithm (82% correct)
- Fuzzy ARTMAP (96% correct)

Circle-in-the-square task
- Back propagation (90% correct)
- Fuzzy ARTMAP (99.5% correct)

Two-spiral task
- Back propagation (10,000 - 20,000 training epochs)
- Fuzzy ARTMAP (1-5 training epochs)

■ **Table 2.** *ARTMAP benchmark studies.*

Choose the globally best answer—In many learning algorithms local minima or less-than-optimal solutions are selected to represent the data as learning proceeds. In fuzzy ARTMAP, at any stage of learning, an input exemplar first selects the category whose top-down expectation provides the globally best match. A top-down expectation thus acts as a prototype for the class of all the input exemplars that its category represents. Before learning self-stabilizes, familiar events gain direct access to the "globally best" category without any search, even if they are interspersed with unfamiliar events that drive hypothesis testing for better matching categories. After learning self-stabilizes, every input directly selects the globally best category without any search.

Calibrate confidence—A confidence measure called *vigilance* calibrates how well an exemplar matches the prototype that it selects. In other words, vigilance measures how well the chosen hypothesis matches the data. If vigilance is low, even poor matches are accepted. Many different exemplars can then be incorporated into one category, so compression and generalization by that category are high. If vigilance is high, then even good matches may be rejected, and hypothesis testing may be initiated to select a new category. In this case, few exemplars activate the same category, so compression and generalization are low. A high level of vigilance can select a unique category for a rare event that predicts an outcome different from that of any of the similar exemplars that surround it.

The Minimax Learning Rule is realized by adjusting the vigilance parameter in response to a predictive error. Vigilance is increased just enough to initiate hypothesis testing to discover a better category, or hypothesis, with which to match the data. In this way, a minimum amount of generalization is sacrificed to correct the error. This process is called match tracking because vigilance tracks the degree of match between exemplar and prototype in response to a predictive error.

Perform rule Extraction—At any stage of learning, a user can translate the state of a fuzzy ARTMAP system into an algorithmic set of rules. From this perspective, fuzzy ARTMAP can be interpreted as a type of self-organizing expert system. These rules evolve as the system is exposed to new inputs. This feature is particularly important in applications such as medical diagnosis from a large database of patient records. Some medical and other benchmark studies that compare the performance of fuzzy ARTMAP with alternative recognition and prediction models are summarized in Table 2. One of the benchmarks is discussed below, and others are described in two references [5-6].

Properties Scale—One of the most serious deficiencies of many artificial intelligence algorithms is that their desirable properties tend to break down as small-scale problems are generalized to large-scale problems. In contrast, all of the desirable properties of fuzzy ARTMAP scale to arbitrarily large problems. However, fuzzy ARTMAP is meant to solve a particular type of problem — it is not intended to solve all problems of learning or intelligence. The categorization and prediction problems that ARTMAP does handle well are core problems in many intelligent systems, and have been technology bottlenecks for many alternative approaches.

A summary is now given of Adaptive Resonance Theory, or ART, networks for unsupervised learning and categorization. Then a connection between certain ART systems and fuzzy logic is noted. Fuzzy ART networks for unsupervised learning and categorization are next described. Finally, fuzzy ART modules are combined into a fuzzy ARTMAP system that is capable of supervised learning, recognition, and prediction. A benchmark comparison of fuzzy ARTMAP with genetic algorithms is then summarized.

A Review of Unsupervised ART Systems

The Adaptive Resonance Theory, or ART, was introduced as a theory of human cognitive information processing [7-8]. The theory has since led to an evolving series of real-time neural network models for unsupervised category learning and pattern recognition. These models are capable of learning stable recognition categories in response to arbitrary input sequences with either fast or slow learning. Model families include ART 1 [9], which can learn to categorize binary input patterns presented in an arbitrary order; ART 2 [10], which can learn to categorize either analog or binary input patterns presented in an arbitrary order; and ART 3 as in [11], which can carry out parallel searches by testing hypotheses about distributed recognition codes in a multilevel network hierarchy. Variations of these models adapted to the demands of individual applications have been developed by a number of authors.

An example from the family of ART 1 models and a typical ART search cycle are illustrated in Figs. 4 and 5, respectively. Level F_1 in Fig. 4 contains a network of nodes, each of which represents a particular combination of sensory features. Level F_2 contains a network of nodes that represent recognition codes that are selectively activated by patterns of activation across F_1. The activities of nodes in F_1 and F_2 are also called short term memory (STM) traces. STM is the type of memory that can be rapidly reset without leaving an enduring trace. For instance, it is easy to reset a person's STM of a list of numbers by distracting the person with an unexpected event. STM is distinct from LTM, or long term memory, which is the type of memory that we usually ascribe to learning. For example, we do not

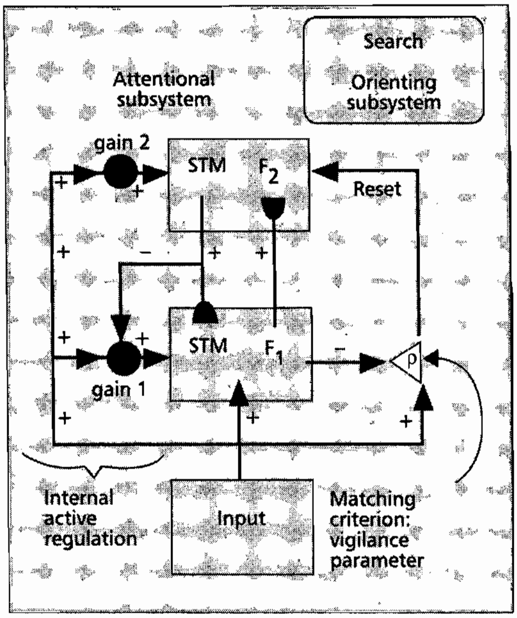


Figure 4. *Typical ART 1 neural network.*

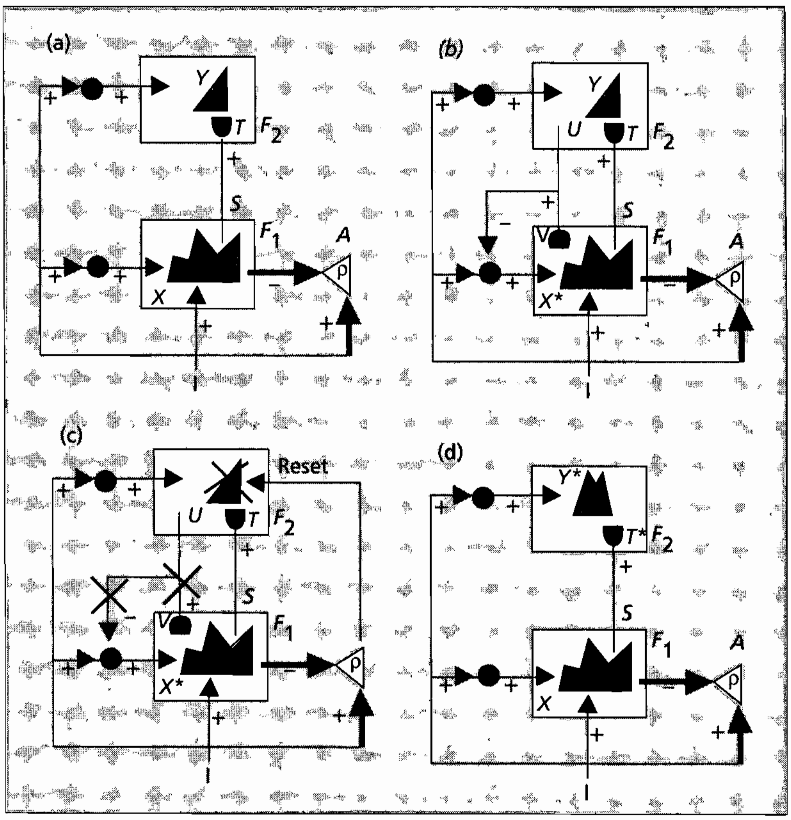


Figure 5. *ART search for an* F_2 *code: (a) The input pattern* ***I*** *generates the specific STM activity pattern* ***X*** *at* F_1 *as it nonspecifically activates the orienting subsystem A. Pattern* ***X*** *both inhibits A and generates the output signal pattern S. Signal pattern S is transformed into the input pattern T, which activates the STM pattern* ***Y*** *across* F_2*. (b) Pattern* ***Y*** *generates the top-down signal pattern U which is transformed into the prototype pattern V. If V mismatches* ***I*** *at* F_1*, then a new STM activity pattern* ***X**** *is generated at* F_1*. The reduction in total STM activity that occurs when* ***X*** *is transformed into* ***X**** *causes a decrease in the total inhibition from* F_1 *to A. (c) If the matching criterion fails to be met, A releases a nonspecific arousal wave to* F_2*, which resets the STM pattern at* F_2*. (d) After* ***Y*** *is inhibited, its top-down prototype signal is eliminated, and* ***X*** *can be reinstated at* F_1*. Enduring traces of the prior reset lead* ***X*** *to activate a different STM pattern* ***Y**** *at* F_2*. If the top-down prototype due to* ***Y**** *also mismatches* ***I*** *at* F_1*, then the search for an appropriate* F_2 *code continues.*

The criterion of an acceptable 2/3 Rule match is defined by a dimensionless parameter called vigilance.

forget our parent's names when we are distracted by an unexpected event.

As shown in Fig. 5a, an input vector **I** registers itself as a pattern **X** of activity across level F_1. The F_1 output vector **S** is then transmitted through the multiple converging and diverging adaptive filter pathways emanating from F_1. This transmission event multiplies vector **S** by a matrix of adaptive weights, or LTM traces, to generate a net input vector **T** to level F_2. The internal competitive dynamics of F_2 contrast-enhance vector **T**. A compressed activity vector **Y** is thereby generated across F_2. In ART 1, the competition is tuned so that the F_2 node that receives the maximal $F_1 \rightarrow F_2$ input is selected. Only one component of **Y** is nonzero after this choice takes place. Activation of such a winner-take-all node defines the category, or symbol, of **I**. Such a category represents all inputs, **I**, that maximally activate the corresponding node. So far, these are the rules of a self-organizing feature map, also called competitive learning or learned vector quantization. Such models were developed by Grossberg [12-16] and von der Malsburg [17-18]. Cohen and Grossberg [19-20], Grossberg and Kuperstein [21], and Kohonen [22] have applied them extensively to problems in speech recognition and adaptive sensory-motor control, among others.

Activation of an F_2 node may be interpreted as "making a hypothesis" about input **I**. When **Y** is activated, it generates an output vector **U** that is sent top-down through the second adaptive filter. After multiplication by the adaptive weight matrix of the top-down filter, a net vector **V** is input to F_1 (Fig. 5b). Vector **V** plays the role of a learned top-down expectation. Activation of **V** by **Y** may be interpreted as "testing the hypothesis" **Y**, or "reading out the category prototype" **V**. The ART 1 network is designed to match the "expected prototype" **V** of the category against the active input pattern, or exemplar, **I**. Nodes that are activated by **I** are suppressed if they do not correspond to large LTM traces in the prototype pattern **V**. Thus F_1 features not "expected" by **V** are suppressed. Expressed in a different way, the matching process may change the F_1 activity pattern **X** by suppressing activation of all the feature detectors in **I** that are not "confirmed" by hypothesis **Y**. The resultant pattern **X*** encodes the cluster of features in **I** that the network deems relevant to the hypothesis **Y**, based upon its past experience. Pattern **X*** encodes the pattern of features to which the network "pays attention."

If **V** is close enough to the input **I**, then a state of resonance develops as the attentional focus takes hold. Pattern **X*** of attended features reactivates hypothesis **Y** which, in turn, reactivates **X***. The network locks into a resonant state through the mutual positive feedback that dynamically links **X*** with **Y**. The resonant state persists long enough for learning to occur; hence the term Adaptive Resonance Theory. ART systems learn prototypes, rather than exemplars, because the attended feature vector **X***, rather than **I** itself, is learned.

This attentive matching process is realized by combining three different types of inputs at level F_1: bottom-up inputs, top-down expectations, and attentional gain control signals (Fig. 4). The attentional gain control channel sends the same signal to all F_1 nodes; it is a "nonspecific," or

ART 1 (binary)	Fuzzy ART (analog)
Category choice	
$T_j = \dfrac{\lvert I \cap w_j \rvert}{\alpha + \lvert w_j \rvert}$	$T_j = \dfrac{\lvert I \wedge w_j \rvert}{\alpha + \lvert w_j \rvert}$
Match criterion	
$\dfrac{\lvert I \cap w \rvert}{\lvert I \rvert} \geq \rho$	$\dfrac{\lvert I \wedge w \rvert}{\lvert I \rvert} \geq \rho$
Fast learning	
$w_j^{(new)} = I \cap w_j^{(old)}$	$w_j^{(new)} = I \wedge w_j^{(old)}$
$\cap$ = logical AND intersection	$\wedge$ = fuzzy AND minimum

■ **Figure 6.** *Comparison of ART 1 and Fuzzy ART.*

modulatory, channel. Attentive matching obeys the 2/3 Rule: an F_1 node can be fully activated only if two of the three input sources that converge upon it send positive signals at a given time [9]. The 2/3 Rule shows how an ART system can be "primed" to expect a subsequent event. A top-down expectation activates to subthreshold levels the F_1 nodes in its prototype. None of the nodes are activated well enough to generate output signals in the absence of their "second third" of the 2/3 Rule. They are nonetheless "primed," or ready, to fire rapidly and vigorously if a bottom-up input does match their prototype well enough. Thus ART systems are "intentional" or "goal-oriented" systems in the sense that their expectations can be primed to selectively seek out data in which they are interested.

The 2/3 Rule also allows an ART system to react to inputs in the absence of prior priming, because a bottom-up input directly activates its target F_1 features and indirectly activates them via the nonspecific gain control channel to satisfy the 2/3 Rule (Fig. 5a). After the input instates itself at F_1, which leads to selection of hypothesis **Y** and top-down expectation **V**, the 2/3 Rule ensures that only those F_1 nodes that are confirmed by the expectation can remain active in STM.

The criterion of an acceptable 2/3 Rule match is defined by a dimensionless parameter called vigilance. Vigilance weighs how close exemplar **I** must be to the top-down prototype **V** in order for resonance to occur. Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization, or morphological variability, can be learned by a single ART system. Low vigilance leads to broad generalization and abstract prototypes. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars. Within the limit of very high vigilance, prototype learning reduces to exemplar learning. Thus, a single ART system may be used, say, to recognize abstract categories of faces and dogs, as well as individual faces and dogs. Exemplars can be coded by specialized "grandmother cells" at the same time that abstract prototypes are coded by general categories of the same network. The particular combination of prototypes that is learned depends upon the predictive success of the learned categories in a particular task environment.

If the top-down expectation **V** and the bottom-up input **I** are too novel, or unexpected, to satisfy the vigilance criterion, then a bout of hypothesis testing — a memory search — is triggered. Searching leads to the selection of a better recognition code, symbol, category, or hypothesis to represent **I** at level F_2. An orienting subsystem mediates the search process (Fig. 4). The orienting subsystem interacts with the attentional subsystem to enable the attentional subsystem to learn new F_2 representations with which to remember novel events without risking unselective forgetting of its previous knowledge (Figs. 5c and 5d).

The search process prevents associations from forming between **Y** and **X*** if **X*** is too different from **I** to satisfy the vigilance criterion. The search process resets **Y** before such an association can form, as shown in Fig. 5c. A familiar category may be selected by the search if its prototype is similar enough to **I** to satisfy the vigilance criterion. The prototype may then be refined in light of new information carried by **I**. If **I** is too different from any of the previously learned prototypes, then an uncommitted F_2 node is selected and learning of a new category is initiated. A network parameter controls how far the search proceeds before an uncommitted node is chosen.

As inputs that correspond to a particular category are practiced over learning trials, the search process converges upon a stable learned recognition category in F_2. This process corresponds to making the inputs "familiar" to the network. After familiarization takes place, all inputs coded by that category access it directly in a single pass, and searching is automatically disengaged. The selected category's prototype provides the globally best match to the input pattern. While stable learning proceeds online, familiar inputs directly activate their categories and novel inputs continue to trigger adaptive searches for better categories, until the network's memory capacity is reached.

Uses for ART Systems

ART systems have been used to explain and predict a variety of cognitive and brain data that have as yet received no other theoretical explanation [23-26]. A formal lesion of the orienting subsystem, for example, creates a memory disturbance that mimics properties of medial temporal amnesia [27-28]. These and related data correspondences to orienting properties have led to a neurobiological interpretation of the orienting subsystem in terms of the hippocampal formation of the brain. In visual object-recognition applications, the interactions within the F_1 and F_2 levels of the attentional subsystem are interpreted in terms of data concerning the prestriate visual cortex and the inferotemporal cortex [29], with the attentional gain control pathway interpreted in terms of the pulvinar region of the brain.

From a computer science perspective, ART systems have an interpretation that is no less interesting. The read-out of top-down expectation **V** may be interpreted as a type of hypothesis-driven query. The matching process at F_1 and the hypothesis testing process at F_2 may be interpreted as query-driven symbolic substitutions. From this perspective, ART systems provide examples of new types of self-organizing production systems [30]. This interpretation of ART networks as production systems indicates how they contribute to artificial intelligence, a major goal of which is to understand the cognitive operations of human thinking in terms of production systems. The ability of ART production systems to explain many cognitive and neurobiological data that cannot be explained by classical production systems illustrates how ART systems have brought us closer to realizing this goal of artificial intelligence.

ARTMAP Systems

By incorporating predictive feedback into their control of the hypothesis testing cycle, the ARTMAP systems that are described below embody self-organizing production systems that are also goal-oriented. The fact that fuzzy logic may also be usefully incorporated into ART systems blurs the traditional boundaries between artificial intelligence and neural networks even further.

ARTMAP systems are capable of compressing different sorts of information into many distinct recognition categories that may all be used to make the same prediction, as shown in Fig. 2. The expertise of such an ARTMAP system can be inferred by a direct study of the rules it uses to arrive at predictions. This may be done at any stage of the learning process.

Suppose, for example, that the input vectors in Fig. 2 are of biochemicals instead of letter fonts, and that the outputs are indices of desired drug effects on behavior rather than letter names. There may be multiple ways in which different biochemicals can achieve the same clinical effect on behavior. At any point in the learning process, the operator of an ARTMAP system can test how many recognition categories have been detected that give rise to the desired clinical effect. The operator simply needs to check which LTM traces are large in the pathways from learned recognition categories to the desired output node. Within each recognition category, the prototype, or vector of large LTM traces, characterizes a particular rule or bundle of biochemical features that predicts the desired clinical effect. The "if-then" nature of the rule derives from the associative nature of ARTMAP predictions: "if the biochemical has features close enough to a particular prototype, then it predicts the desired outcome." A list of all the prototype vectors provides a transparent set of rules whereby one can predict the desired outcome.

Many such rules may coexist without mutual interference due to the competitive interactions whereby each hypothesis **Y** in Fig. 5 is compressed. Associative networks such as back propagation often mix multiple rules with the same LTM traces because they do not have the competitive dynamics to separate them.

This particular type of rule-based system may also exhibit aspects of "creativity." ARTMAP systems, albeit "supervised," do not use the correct answers to directly force changes in LTM traces, as do supervised systems such as back propagation. ARTMAP systems use the fact that its answers are wrong, along with its present state of knowledge, to test new hypotheses until it discovers, on its own, new representations that are capable of predicting the correct answers.

ART systems have been used to explain and predict a variety of cognitive and brain data that have as yet received no other theoretical explanation.

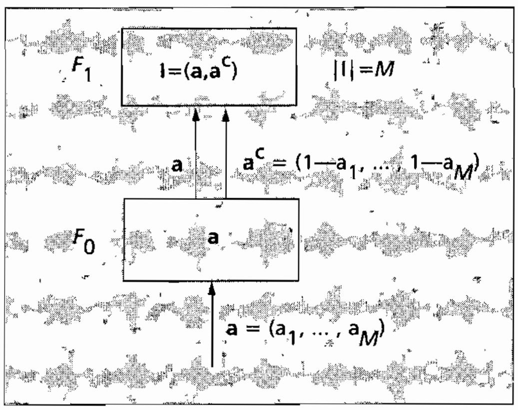
Fuzzy ARTMAP has been benchmarked against a variety of machine learning, neural network, and genetic algorithms with considerable success.

ART Systems and Fuzzy Logic

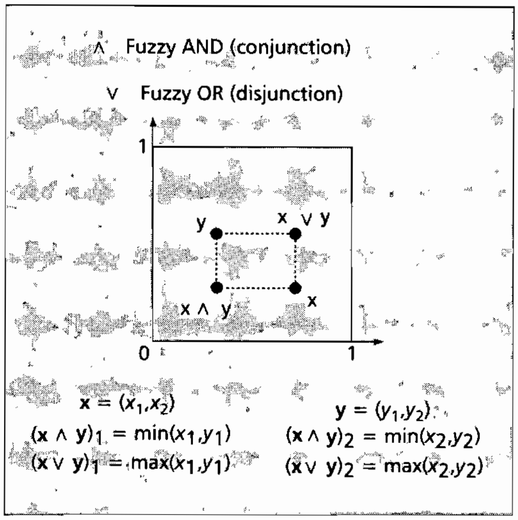
Fuzzy ART is a form of ART 1 that incorporates fuzzy logic operations [31]. Although ART 1 can learn to classify only binary input patterns, fuzzy ART can learn to classify both analog and binary input patterns. In addition, fuzzy ART reduces to ART 1 in response to binary input patterns. Learning both analog and binary input patterns is achieved by replacing appearances of the intersection operator in ART 1 by the MIN operator of fuzzy set theory (Fig. 6). The MIN operator ($\wedge$) reduces to the intersection operator ($\cap$) in the binary case. Of particular interest is the fact that, as parameter α approaches zero, function Tj, which controls category choice through the bottom-up filter (Fig. 5), reduces to the operation of fuzzy subsethood [32]. T_j then measures the degree to which the adaptive weight vector $\mathbf{w}_j$ is a fuzzy subset of input vector $\mathbf{I}$.

In fuzzy ART, input vectors are normalized at a preprocessing stage (Fig. 7). This normalization procedure, called complement coding, leads to a symmetric theory in which the MIN operator ($\wedge$) and the MAX operator ($\vee$) of fuzzy set theory play complementary roles [33]. The categories formed by fuzzy ART are then hyper-rectangles. Fig. 8 illustrates how MIN and MAX define these rectangles in the 2-dimensional case. The MIN and MAX values define the acceptable range of feature variation in each dimension. Complement coding uses on-cells (with activity $\mathbf{a}$ in Fig. 7) and off-cells (with activity $\mathbf{a}^c$) to represent the input pattern, and preserves individual feature amplitudes while normalizing the total on-cell/off-cell vector.

The on-cell portion of a prototype encodes features that are critically present in category exemplars, while the off-cell portion encodes features that are critically absent. Each category is then defined by an interval of expected values for each input feature. For instance, we learn by example that men usually have hair on their heads. Fuzzy ART would encode this feature a wide interval ([A, 1]) of expectations of "hair on head" for the category "man". Similarly, since men sometimes wear hats, the feature "hat on head" would be encoded by a wide interval ([0, B]) of expectations. On the other hand, a dog almost always has hair on its head but almost never wears a hat. These features for the category "dog" would thus be encoded by two narrow intervals ([C, 1] for hair and [0, D] for hat) corresponding to narrower ranges of expectations for these two features.



■ **Figure 7.** *Complement coding uses on-cell and off-cell pairs to normalize input vectors.*



■ **Figure 8.** *Fuzzy AND and OR operations generate category hyper-rectangles.*

Learning in fuzzy ART is stable because all adaptive weights can only decrease in time. Decreasing weights correspond to increasing sizes of category "boxes". Smaller vigilance values lead to larger category boxes, and learning stops when the input space is covered by boxes. The use of complement coding works with the property of increasing box size to prevent a proliferation of categories. With fast learning, constant vigilance, and a finite input set of arbitrary size and composition, learning stabilizes after just one presentation of each input pattern. A fast-commit, slow-recode option combines fast learning with a forgetting rule that buffers system memory against noise. Using this option, rare events can be rapidly learned, yet previously learned memories are not rapidly erased in response to statistically unreliable input fluctuations. See the appendix entitled "Fuzzy ART Algorithm" for an explanation of defining equations of fuzzy ART.

When the supervised learning of fuzzy ARTMAP controls category formation, a predictive error can force the creation of new categories that could not otherwise be learned due to monotone increases in category size through time in the unsupervised case. Supervision permits the creation of complex categorical structures without a loss of stability.

Fuzzy ARTMAP

Each fuzzy ARTMAP system includes a pair of fuzzy ART modules (ART$_a$ and ART$_b$), as shown in Fig. 9. During supervised learning, ART$_a$ receives a stream $\{\mathbf{a}^{(p)}\}$ of input patterns and ART$_b$ receives a stream $\{\mathbf{b}^{(p)}\}$ of input patterns, where $\mathbf{b}^{(p)}$ is the correct prediction given $\mathbf{a}^{(p)}$. These modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time.

The controller is designed to create the minimal number of ART$_a$ recognition categories, or "hidden units," needed to meet accuracy criteria. As noted above, this is accomplished by realizing a Minimax Learning Rule that conjointly minimizes predictive error and maximizes predictive generalization. This scheme automatically links predictive success to category size on a trial-by-trial basis using only local operations. It works by increasing the vigilance parameter ρ_a of ART$_a$ by the minimal amount needed to correct a predictive error at ART$_b$ (Fig. 10).

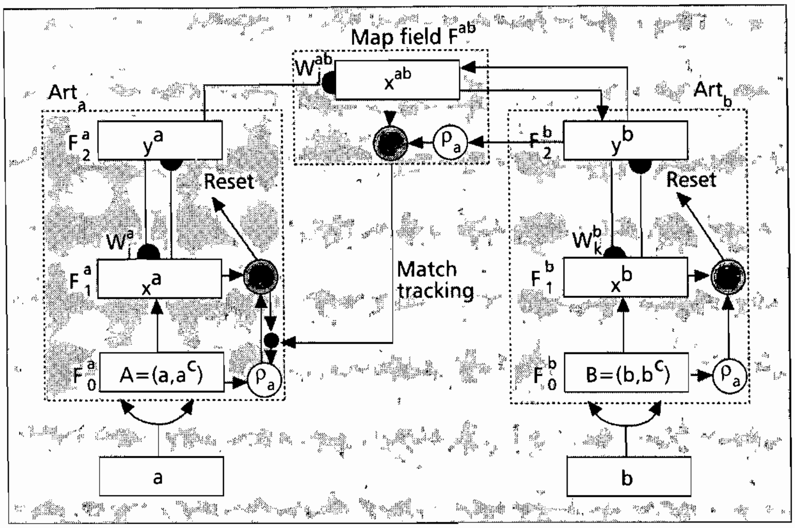
Parameter ρ_a calibrates the minimum confidence that ART$_a$ must have in a recognition category (hypothesis) that is activated by an input $\mathbf{a}^{(p)}$, in order for ART$_a$ to accept that category instead of searching for a better one through an automatically controlled process of hypothesis testing. As in ART 1, lower values of ρ_a enable larger categories to form. These lower ρ_a values lead to broader generalization and higher code compression. A predictive failure at ART$_b$ increases the minimal confidence ρ_a by the least amount needed to trigger hypothesis testing at ART$_a$, using a mechanism called match tracking [5]. *Match tracking* sacrifices the minimum amount of generalization necessary to correct the predictive error.

Match tracking presents the idea that the system must have accepted hypotheses with too little confidence to satisfy the demands of a particular environment; it increases the criterion confidence just enough to trigger hypothesis testing. Hypothesis testing leads to the selection of a new ART$_a$ category, which focuses attention on a new cluster of $\mathbf{a}^{(p)}$ input features that is better able to predict $\mathbf{b}^{(p)}$. Due to the combination of match tracking and fast learning, a single ARTMAP system can learn a different prediction for a rare event than for a cloud of similar frequent events in which it is embedded. The equations for fuzzy ART and fuzzy ARTMAP are given in the appendix in algorithmic form.

A Fuzzy ARTMAP Benchmark

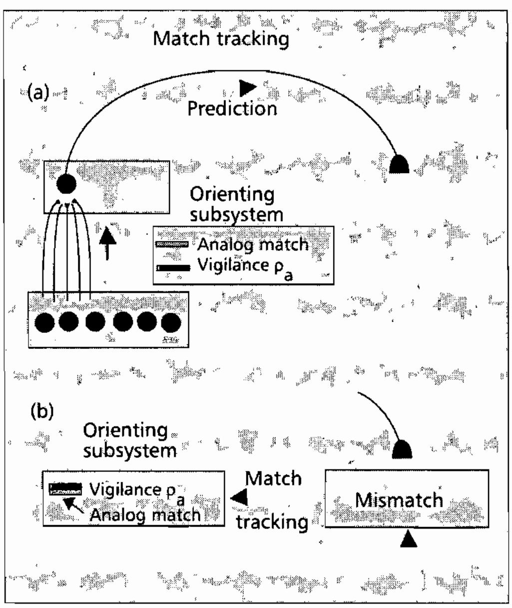
Fuzzy ARTMAP has been benchmarked against a variety of machine learning, neural network, and genetic algorithms with considerable success (Table 2). One study used a benchmark machine learning task that Frey and Slate developed and described as a "difficult categorization problem" [34]. The task requires a system to identify an input exemplar as one of 26 capital letters, A through Z. The database was derived from 20,000 unique black-and-white pixel images. The task is difficult because of the wide variety of letter types represented: The twenty "fonts represent five different stroke styles (simplex, duplex, complex, and Gothic) and six different letter styles (block, script, italic, English, Italian, and German)." In addition, each image was randomly distorted, leaving many of the characters misshapen. Sixteen numerical feature attributes were then obtained from each character image, and each attribute value was scaled to a range of 0 to 15. The resulting Letter Image Recognition file is archived in the UCI Repository of Machine Learning Databases and Domain Theories, maintained by David Aha and Patrick Murphy (ml_repository@ics.uci.edu on Internet).

Frey and Slate used this database to test performance of a family of classifiers based on Holland's genetic algorithms [35]. The training set consisted of 16,000 exemplars, with the remaining 4000 exemplars used for testing. Genetic algorithm classifiers having different input representations, weight-update and rule-creation schemes, and system parameters were systematically compared.



■ **Figure 9.** *Fuzzy ARTMAP architecture. The ART$_a$ complement coding preprocessor transforms the M_a-vector **a** into the $2M_a$-vector $A = (\mathbf{a},\mathbf{a}^c)$ at the ART$_a$ field F_0^a. A is the input vector to the ART$_a$ field F_1^a. Similarly, the input to F_1^b is the $2M_b$-vector $(\mathbf{b},\mathbf{b}^c)$. When a prediction by ART$_a$ is disconfirmed at ART$_b$, inhibition of map field activation induces the match tracking process. Match tracking raises the ART$_a$ vigilance ρ_a to just above the F_1^a to F_0^a match ratio $|\mathbf{x}^a|/|\mathbf{A}|$. This triggers an ART$_a$ search which leads to activation of either an ART$_a$ category that correctly predicts **b** or to a previously uncommitted ART$_a$ category node.*

Training was carried out for five epochs, plus a sixth "verification" pass during which no new rules were created but a large number of unsatisfactory rules were discarded. In Frey and Slate's comparative study, these systems had correct prediction rates that ranged from 24.5 percent to 80.8 percent on the 4000-item test set. The best performance was obtained using an integer input representation, a reward-sharing weight update, an exemplar method of rule creation, and a parameter setting that allowed an unused or erroneous rule to stay in



■ **Figure 10.** *Match tracking: (a) A prediction is made by ART$_a$ when the vigilance ρ_a is less than the analog match value. (b) A predictive error at ART$_b$ increases the vigilance value of ART$_a$ until it just exceeds the analog match value, and thereby triggers hypothesis testing that searches for a more predictive bundle of features to which to attend.*

Unsupervised ART modules have found their way into diverse applications.

the system for a long time before being discarded. After training in the best case, 1302 rules and 8 attributes per rule were created, as were over 35,000 more rules that were discarded during verification. (For purposes of comparison, a rule is somewhat analogous to an ART_a category in ARTMAP, and the number of attributes per rule is analogous to the size $|\mathbf{w}_j^a|$ of an ART_a category weight vector.)

Building on the results of their comparative study, Frey and Slate investigated two types of alternative algorithms: an accuracy-utility bidding system that had slightly improved performance (81.6 percent) in the best case, and an exemplar/hybrid rule creation scheme that further improved performance to a maximum of 82.7 percent but required the creation of over 100,000 rules prior to the verification step.

Fuzzy ARTMAP had an error rate on the letter-recognition task that was consistently less than one third that of the three best Frey-Slate genetic algorithm classifiers described above. In particular, after one to five epochs, individual fuzzy ARTMAP systems had a robust prediction rate of 90 to 94 percent on the 4000-item test set. A voting strategy consistently improved this performance. The voting strategy is based on the observation that ARTMAP fast learning typically leads to different adaptive weights and recognition categories for different orderings of a given training set, even when overall predictive accuracy of all simulations is similar.

The different category structures cause the set of test items where errors occur to vary from one simulation to the next. The voting strategy uses an ARTMAP system that is trained several times on input sets with different orderings. The final prediction for a given test set item is the one made by the largest number of simulations. Because the set of items making erroneous predictions varies from one simulation to the next, voting cancels many of the errors.

Such a voting strategy can also be used to assign confidence estimates to competing predictions given small, noisy, or incomplete training sets. Voting consistently eliminated 25 to 43 percent of the errors, giving a robust prediction rate of 92 to 96 percent. Fuzzy ARTMAP simulations each created fewer than 1070 ART_a categories, compared to the 1040 to 1302 final rules of the three genetic classifiers with the best performance rates. Most fuzzy ARTMAP learning occurred on the first epoch, with test set performance on systems trained for one epoch typically over 97 percent of that of systems exposed to inputs for five epochs.

Conclusion

Fuzzy ARTMAP is one of a rapidly growing family of attentive self-organizing learning, hypothesis testing, and prediction systems that have evolved from the biological theory of cognitive information processing of which ART forms an important part [16, 23-26]. Unsupervised ART modules have found their way into such diverse applications as the control of mobile robots, learning and searching of airplane part inventories, medical diagnosis, 3-D visual object recognition, music recognition, seismic recognition, sonar recognition, and laser radar recognition [36-40].

All of these applications exploit the ability of ART systems to rapidly learn to classify large databases in a stable fashion, to calibrate their confidence in a classification, and to focus attention upon those groups of features that they deem to be important based upon their past experience. We anticipate that the growing family of supervised ARTMAP systems will find an even broader range of applications due to their ability to adapt the number, shape, and scale of their category boundaries to meet the online demands of large nonstationary databases.

The algorithmic equations that define fuzzy ART and fuzzy ARTMAP are summarized in the appendix that follows.

Appendix
Fuzzy ART Algorithms

Proliferation of categories is avoided in fuzzy ART if inputs are normalized.

ART field activity vectors—Each ART system includes a field F_0 of nodes that represent a current input vector; a field F_1 that receives both bottom-up input from F_0 and top-down input from a field F_2 that represents the active code, or category. The F_0 activity vector is denoted $\mathbf{I} = (I_1, \ldots I_M)$, with each component I_i in the interval $[0,1]$, $i = 1, \ldots, M$. The F_1 activity vector is denoted $\mathbf{x} = (x_1, \ldots, x_M)$ and the F_2 activity vector is denoted $\mathbf{y} = (y_1, \ldots, y_N)$. The number of nodes in each field is arbitrary.

Weight vector—Associated with each F_2 category node j $(j = 1, \ldots, N)$ is a vector $\mathbf{w}_j \equiv (w_{j1}, \ldots, w_{jM})$ of adaptive weights, or LTM traces. Initially,

$$w_{j1}(0) = \ldots = w_{jM}(0) = 1; \tag{1}$$

Each category is then said to be uncommitted. After a category is selected for coding, it becomes committed. As shown below, each LTM trace w_{ji} is monotone nonincreasing through time and hence converges to a limit. The fuzzy ART weight vector $\mathbf{w}_j$ subsumes both the bottom-up and top-down weight vectors of ART 1.

Parameters—Fuzzy ART dynamics are determined by a choice parameter $\alpha > 0$; a learning rate parameter $\beta \in [0,1]$; and a vigilance parameter $\rho \in [0,1]$.

Category choice—For each input $\mathbf{I}$ and F_2 node j, the choice function T_j is defined by:

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{2}$$

where the fuzzy AND [33] operator $\wedge$ is defined by:

$$(\mathbf{p} \wedge \mathbf{q})_i \equiv \min(p_i, q_i) \tag{3}$$

and where the norm $|\cdot|$ is defined by:

$$|\mathbf{p}| \equiv \sum_{i=1}^{M} |p_i|. \tag{4}$$

for any M-dimensional vectors $\mathbf{p}$ and $\mathbf{q}$. For notational simplicity, $T_j(\mathbf{I})$ in (2) is often written as T_j when the input $\mathbf{I}$ is fixed.

The system is said to make a category choice when at most one F_2 node can become active at a given time. The category choice is indexed by J, where

$$T_J = \max\{T_j : j = 1 \ldots N\}. \tag{5}$$

If more than one T_J is maximal, the category j with the smallest index is chosen. In particular, nodes become committed in order $j = 1, 2, 3, \ldots$. When the J^{th} category is chosen, $y_J = 1$; and $y_j = 0$ for $j \neq J$. In a choice system, the F_1 activity vector x obeys the equation

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{if } F_2 \text{ is inactive} \\ \mathbf{I} \wedge \mathbf{w}_J & \text{if the } J^{th}\ F_2 \text{ node is chosen .} \end{cases} \tag{6}$$

Resonance or reset—Resonance occurs if the match function $|\mathbf{I} \wedge \mathbf{w}_J| / |\mathbf{I}|$ of the chosen category meets the vigilance criterion:

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho; \tag{7}$$

that is, by (6), when the J^{th} category is chosen, resonance occurs if

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| \geq \rho |\mathbf{I}|. \tag{8}$$

Learning then ensues, as defined later in this sidebar. Mismatch reset occurs if

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} < \rho; \tag{9}$$

that is, if:

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| < \rho |\mathbf{I}|. \tag{10}$$

Then the value of the choice function T_J is set to 0 for the duration of the input presentation to prevent the persistent selection of the same category during search. A new index J is then chosen, by (5). The search process continues until the chosen J satisfies (7).

Learning—Once search ends, the weight vector $\mathbf{w}_J$ is updated according to the equation

$$\mathbf{w}_J^{(new)} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{(old)}) + (1 - \beta)\mathbf{w}_J^{(old)}. \tag{11}$$

Fast learning corresponds to setting $\beta = 1$. The learning law used in the EACH system of Salzberg is equivalent to equation (11) in the fast-learn limit with the complement coding option described below [41].

Fast-commit slow-recode option—For efficient coding of noisy input sets, it is useful to set $\beta = 1$ when J is an uncommitted node, and then to take $\beta < 1$ after the category is committed. Then $\mathbf{w}_J^{(new)} = \mathbf{I}$, and the first time category J becomes active. Moore introduced the learning law (11), with fast commitment and slow recoding, to investigate a variety of generalized ART 1 models [42]. Some of these models are similar to fuzzy ART, but none includes the complement coding option. Moore described a category proliferation problem that can occur in some analog ART systems when a large number of inputs erode the norm of weight vectors. Complement coding solves this problem.

Input normalization/complement coding option—Proliferation of categories is avoided in fuzzy ART if inputs are normalized. Complement coding is a normalization rule that preserves amplitude information. Complement coding represents both the on-response and the off-response to an input vector $\mathbf{a}$ (Fig. 7). To define this operation in its simplest form, let $\mathbf{a}$ itself represent the on-response. The complement of $\mathbf{a}$, denoted

The fuzzy ARTMAP system incorporates two fuzzy ART linked together via an inter-ART module F^{ab} called a map field.

by $\mathbf{a}^c$, represents the off-response, where

$$a_i^c \equiv 1 - a_i. \tag{12}$$

The complement-coded input **I** to the field F1 is the 2M-dimensional vector:

$$\mathbf{I} = (\mathrm{a}, \mathrm{a}^c) \equiv (a_1, \ldots, a_M, a_1^c, \ldots, a_M^c). \tag{13}$$

Note that

$$\begin{aligned} |\mathbf{I}| &= \left|\left(\mathbf{a}, \mathbf{a}^c\right)\right| \\ &= \sum_{i=1}^{M} a_i + \left(M - \sum_{i=1}^{M} a_i\right) \\ &= M, \end{aligned} \tag{14}$$

so inputs preprocessed into complement coding form are automatically normalized. Where complement coding is used, the initial condition (1) is replaced by

$$w_{j1}(0) = \ldots = w_{j,2M}(0) = 1. \tag{15}$$

Fuzzy ARTMAP Algorithm

The fuzzy ARTMAP system incorporates two fuzzy ART, modules ART_a and ART_b, that are linked together via an inter-ART module F^{ab} called a map field. The map field is used to form predictive associations between categories and to realize the match tracking rule whereby the vigilance parameter of ART_a increases in response to a predictive mismatch at ART_b. The interactions mediated by the map field F^{ab} may be operationally characterized as follows.

ART_a and ART_b—Inputs to ART_a and ART_b are in the complement code form: for ART_a, $\mathbf{I} = \mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$; for ART_b, $\mathbf{I} = \mathbf{B} = (\mathbf{b}, \mathbf{b}^c)$ (Fig. 9). Variables in ART_a or ART_b are designated by subscripts or superscripts "a" or "b". For ART_a, let $\mathbf{x}^a \equiv (x_1^a \ldots x_{2Ma}^a)$ denote the F_1^a output vector; let $\mathbf{y}^a \equiv (y_1^a \ldots y_{Na}^a)$ denote the F_2^a output vector; and let $\mathbf{w}_j^a \equiv (w_{j1}^a, w_{j2}^a, \ldots, w_{j,2M_a})$ denote the j^{th} ART_a weight vector. For ART_b, let $\mathbf{x}^b \equiv (x_1^b \ldots x_{2M_b}^b)$ denote the F_1^b output vector; let $\mathbf{y}^b \equiv (y_1^b \ldots y_{Nb}^b)$ denote the F_2^b output vector; and let $\mathbf{w}_k^b \equiv (w_{k1}^b, w_{k2}^b, \ldots, w_{k,2M_b}^b)$ denote the k^{th} ART_b weight vector. For the map field, let $\mathbf{x}^{ab} \equiv (x_1^{ab}, \ldots, x_{N_b}^{ab})$ denote the F^{ab} output vector, and let $\mathbf{w}_j^{ab} \equiv (w_{j1}^{ab}, \ldots, w_{jN_b}^{ab})$ denote the weight vector from the j^{th} F_2^a node to F^{ab}. Vectors $\mathbf{x}^a$, $\mathbf{y}^a$, $\mathbf{x}^b$, $\mathbf{y}^b$, and $\mathbf{x}^{ab}$ are set to **0** between input presentations.

Map field activation—The map field F^{ab} is activated whenever one of the ART_a or ART_b categories is active. If node J of F_2^a is chosen, then its weights $\mathbf{w}_J^{ab}$ activate F^{ab}. If node K in F_2^b is active, then the node K in F^{ab} is activated by 1-to-1 pathways between F_2^b and F^{ab}. If both ART_a and ART_b are active, then F^{ab} becomes active only if ART_a predicts the same category as ART_b via the weights $\mathbf{w}_J^{ab}$. The F^{ab} output vector $\mathbf{x}^{ab}$ obeys

$$\mathbf{x}^{ab} = \begin{cases} \mathbf{y}^b \wedge \mathbf{w}_J^{ab} & \text{if the Jth } F_2^a \text{ node is active and } F_2^b \text{ is active} \\ \mathbf{w}_J^{ab} & \text{if the Jth } F_2^a \text{ node is active and } F_2^b \text{ is inactive} \\ \mathbf{y}^b & \text{if } F_2^a \text{ is inactive and } F_2^b \text{ is active} \\ \mathbf{0} & \text{if } F_2^a \text{ is inactive and } F_2^b \text{ is inactive} \end{cases} \tag{16}$$

By (16), $\mathbf{x}^{ab} = 0$ if the prediction $\mathbf{w}_J^{ab}$ is disconfirmed by $\mathbf{y}^b$. Such a mismatch event triggers an ART_a search for a better category.

Match tracking—At the start of each input presentation the ART_a vigilance parameter ρ_a equals a baseline vigilance $\overline{\rho_a}$. The map field vigilance parameter is ρ_{ab}. If

$$|\mathbf{x}^{ab}| < \rho_{ab}|\mathbf{y}^b|, \tag{17}$$

then ρ_a is increased until it is slightly larger than $|\mathbf{A} \wedge \mathbf{w}_J^a||\mathbf{A}|^{-1}$, where $^{\mathbf{A}}$ is the input to F_1^a, in complement coding form. Then,

$$|\mathbf{x}^a| = |\mathbf{A} \wedge \mathbf{w}_J^a| < \rho_a|\mathbf{A}|, \tag{18}$$

where J is the index of the active F_2^a node, as in (10). When this occurs, ART_a search leads either to activation of another F_2^a node J with

$$|\mathbf{x}^a| = |\mathbf{A} \wedge \mathbf{w}_J^a| \geq \rho_a|\mathbf{A}| \tag{19}$$

and

$$|\mathbf{x}^{ab}| = |\mathbf{y}^b \wedge \mathbf{w}_J^{ab}| \geq \rho_{ab}|\mathbf{y}^b|; \tag{20}$$

or, if no such node exists, to the shut-down of F_2^a for the remainder of the input presentation.

Map field learning—Learning rules determine how the map field weights w_{jk}^{ab} change through time, as follows: Weights w_{jk}^{ab} in $\mathrm{F}_2^a \to \mathrm{F}^{ab}$ paths initially satisfy

$$w_{jk}^{ab}(0) = 1. \tag{21}$$

During resonance with the ART_a category J active, $\mathbf{w}_J^{ab}$ approaches the map field vector $\mathbf{x}^{ab}$. With fast learning, once J learns to predict the ART_b category K, that association is permanent; that is, $w_{JK}^{ab} = 1$ for all time.

Acknowledgements

The authors wish to thank Kelly Dumont and Diana Meyers for their valuable assistance in the preparation of the manuscript.

References

[1] D.B. Parker, "Learning-Logic," Invention Report, 581-64, File 1, Office of Technology Licensing, Stanford University, Stanford, Calif., October, 1982.

[2] D.E. Rumelhart, G. Hinton, and R. Williams, "Learning Internal Representations by Error Propagation," Parallel Distributed Processing, D.E. Rumelhart and J.L. McClelland (Eds.), Cambridge, Mass, MIT Press, 1986.

[3] P. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. Thesis, Harvard University, Cambridge, Mass, 1974.

[4] R. Ratcliff, "Connectionist Models of Recognition Memory: Constraints Imposed by Learning and Forgetting Functions," *Psychological Review*, 97, pp. 285-308, 1990.

[5] G.A. Carpenter, S. Grossberg, and J.H. Reynolds, "ARTMAP: Supervised Real-time Learning and Classification of Nonstationary Data by a Self-organizing Neural Network," Neural Networks, 4, pp. 565-88, 1991.

[6] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps," *IEEE Trans. on Neural Networks*, 3, in press, 1992.

[7] S. Grossberg, "Adaptive Pattern Classification and Universal Recoding, II: Feedback, Expectation, Olfaction, and Illusions," *Biological Cybernetics*, 23, pp. 187-202, 1976.

[8] S. Grossberg, "How Does a Brain Build a Cognitive Code?" *Psychological Review*, 1, pp. 1-51, 1980.

[9] G. A. Carpenter and S. Grossberg, "A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine," *Computer Vision, Graphics, and Image Processing*, 37, pp. 54-115, 1987.

[10] G.A. Carpenter and S. Grossberg, "ART2: Stable Self-organization of Pattern Recognition Codes for Analog Input Patterns," *Applied Optics*, 26, pp. 4919-30, 1987.

[11] G.A. Carpenter and S. Grossberg, "ART3: Hierarchical Search Using Chemical Transmitters in Self-organizing Pattern Recognition Architectures," *Neural Networks*, 3, pp.129-152, 1990.

[12] S. Grossberg, "Neural Expectation: Cerebellar and Retinal Analogs of Cells Fired by Learnable or Unlearned Pattern Classes," *Kybernetik*, 10, pp. 49-57, 1972.

[13] S. Grossberg, "On the Development of Feature Detectors in the Visual Cortex with Applications to Learning and Reaction-diffusion Systems," *Biological Cybernetics*, 21, pp. 145-59, 1976.

[14] S. Grossberg, "Adaptive Pattern Classification and Universal Recoding, I: Parallel Development and Coding of Neural Feature Detectors," *Biological Cybernetics*, 23, pp. 121-34, 1976.

[15] S. Grossberg, "A Theory of Human Memory: Self-organization and Performance of Sensory-motor Codes, Maps, and Plans. Progress in Theoretical Biology, Vol. 5., R. Rosen and F. Snell (Eds.), New York, Academic Press, 1978. (Reprinted in S. Grossberg, Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control, Reidel Press, 1982)).

[16] S. Grossberg, Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control, (Reidel Press, 1982).

[17] C. von der Malsburg, "Self-organization of Orientation Sensitive Cells in the Striate Cortex," *Kybernetik*, 14, pp. 85-100, 1973.

[18] D.J. Willshaw and C. von der Malsburg, "How Patterned Neural Connections Can Be Set Up by Self-organization," *Proceedings of the Royal Society of London (B)*, 194. pp. 431-45, 1976.

[19] M. Cohen and S. Grossberg, "Neural Dynamics of Speech and Language Coding: Developmental Programs, Perceptual Grouping, and Competition for Short Term Memory," *Human Neurobiology*, 5, pp. 1-22, 1986.

[20] M. Cohen and S. Grossberg, "Masking Fields: A Massively Parallel Architecture for Learning, Recognizing, and Predicting Multiple Groupings of Patterned Data," *Applied Optics*, 26, pp. 1866-1891, 1987.

[21] S. Grossberg and M. Kuperstein, Neural dynamics of adaptive sensory-motor control: Ballistic eye movements, (Amsterdam, Elsevier/North Holland, 1986). Also in Expanded Edition, pub. by Pergamon Press, 1989).

[22] T. Kohonen, Self-organization and Associative Memory, Springer-Verlag, 1988.

[23] G.A. Carpenter and S. Grossberg (Eds.), Pattern Recognition by Self-organizing Neural Networks, (MIT Press, 1991).

[24] S. Grossberg (Ed.), The Adaptive Brain, I: Cognition, Learning, Reinforcement, and Rhythm, Elsevier/North-Holland, 1987.

[25] S. Grossberg (Ed.), The Adaptive Brain, II: Vision, Speech, Language, and Motor Control, Elsevier/North-Holland, 1987.

[26] S. Grossberg (Ed.), Neural Networks and Natural Intelligence, MIT Press, 1988).

[27] G. Carpenter and S. Grossberg, "Neural Dynamics of Category Learning and Recognition: Attention, Memory Consolidation, and Amnesia" In S. Grossberg (Ed.), The Adaptive Brain, I: Cognition, Learning, Reinforcement, and Rhythm, pp. 238-86, Elsevier/North Holland, 1987.

[28] S. Grossberg and J.W.L. Merrill, "A Neural Network Model of Adaptively Timed Reinforcement Learning and Hippocampal Dynamics," *Cognitive Brain Research*, in press, 1992.

[29] R. Desimone, "Neural Circuits for Visual Attention in the Primate Brain," Neural networks for vision and image processing, G.A. Carpenter and S. Grossberg (Eds.), pp. 343-364, (Cambridge, Mass: MIT Press,1992).

[30] J.E. Laird, A. Newell, and P.S. Rosenbloom, "SOAR: An Architecture for General Intelligence," *Artificial Intelligence*, 33, pp. 1-64, 1987.

[31] G.A. Carpenter, S. Grossberg, and D.B. Rosen, "Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System," *Neural Networks*, 4, pp. 759-71, 1991.

[32] B. Kosko, "Fuzzy Entropy and Conditioning," *Information Sciences*, 40, pp. 165-174, 1986.

[33] L. Zadeh, "Fuzzy Sets," *Information Control*, 8, pp. 338-353, 1965.

[34] P.W. Frey and D.J. Slate, "Letter Recognition Using Holland-style Adaptive Classifiers," *Machine Learning*, 6, pp. 161-82, 1991.

[35] J.H. Holland, "Adaptive Algorithms for Discovering and Using General Patterns in Growing Knowledge Bases," *International Journal of Policy Analysis and Information Systems*, 4, pp. 217-40, 1980.

[36] A.J. Baloch and A.M. Waxman, "Visual Learning, Adaptive Expectations, and Learning Behavioral Conditioning of the Mobil Robot MAVIN," *Neural Networks*, 4, pp. 271-302, 1991.

[37] T. Caudell, S. Smith, C. Johnson, D. Wunsch, and R. Escobedo, "An Industrial Application of Neural Networks to Reusable Design," Adaptive neural systems, Technical Report BCS-CS-ACS-91-001, pp. 185-90, The Boeing Company, Seattle, Wash., 1991).

[38] R.O. Gjerdingen, "Categorization of Musical Patterns by Self-organizing Neuronlike Networks," *Music Perception*, 7, pp. 339-70, 1990.

[39] P. Goodman, V. Kaburlasos, D. Egbert, G. Carpenter, S. Grossberg, J. Reynolds, K. Hammermeister, G. Marshall, and F. Grover, "Fuzzy ARTMAP Neural Network Prediction of Heart Surgery Mortality," Proceedings of the Wang Institute Research Conference: Neural Networks for Learning, Recognition, and Control, p. 48 (Boston, Mass: Boston University, 1992).

[40] M. Seibert and A.M. Waxman, "Learning and Recognizing 3D Objects from Multiple Views in a Neural System," Neural Networks for Perception, H. Wechler (Ed.) Vol. 1, pp.426-444, (New York: Academic Press, 1992).

[41] S.L. Salzberg, Learning with Nested Generalized Exemplars (Boston: Kluwer Academic Publishers, 1990).

[42] B. Moore, "ART 1 and Pattern Clustering," *Proceedings of the 1988 Connectionist Models Summer School*, D. Touretzky, G. Hinton, and T. Sejnowski, Eds., pp. 174-85, Morgan Kaufmann, 1989.

Biographies

GAIL CARPENTER is professor of Cognitive and Neural Systems (CNS) and Mathematics at Boston University. She is the CNS director of graduate studies and a member of the editorial boards of Neural Networks, Neural Computation, and Brain Research. Since receiving a Ph.D. in mathematics (1974), she has carried out research on neural models of vision, nerve impulse generation (Hodgkin-Huxley equations), and complex biological rhythms. Dr. Carpenter's recent work has focused on the development of the adaptive resonance theory architectures (ART 1, 2, and 3) for self-organizing category learning and pattern recognition; and on systems that incorporate ART modules into neural network architectures for incremental supervised learning of binary patterns (ARTMAP) and analog patterns (fuzzy ARTMAP). Her work is supported in part by British Petroleum (89-A-1204), DARPA (AFOSR 90-0083), the National Science Foundation (NSF IRI 90-00530) and the Office of Naval Research (ONR N00014-91-J-4100).

STEPHEN GROSSBERG is Wang Professor of Cognitive and Neural Systems and Professor of Mathematics, Psychology, and Biomedical Engineering at Boston University, where he is the founder and Director of the Center for Adaptive Systems, as well as the founder and Chairman of the Department of Cognitive and Neural Systems. He also organized the Boston Consortium for Behavioral and Neural Studies, which includes investigators from six Boston-area institutions. He founded and was first president of the International Neural Network Society, and is co-editor-in-chief of the society's journal, Neural Networks. Grossberg was General Chairman of the first IEEE International Conference on Neural Networks and was awarded the 1991 IEEE Neural Network Pioneer prize. His work is supported in part by DARPA (AFOSR 90-0083) and the Office of Naval Research (ONR N00014-91-J-4100).

Supervised ARTMAP systems will find a broader range of applications due to their ability to adapt the number, shape, and scale of their category boundaries.